19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011
http://mssanz.org.au/modsim2011

Ultrasound based computer aided diagnosis of breast cancer: Evaluation of a new feature of mass central regularity degree

Ali Al-yousef, Sandhya Samarasinghe

Centre for Advanced Computational Solutions (C-fACS), Lincoln University, Canterbury, New Zealand
Email:ali.al-yousef@lincolnuni.ac.nz

Abstract. Breast cancer is the second leading cause of death after lung cancer in women all over the world. The survival rate of breast cancer patients depends on the stage of diagnosis; patients with stage 0 are more likely to reach cancer free state. Therefore, early detection of breast cancer is the key to patient survival. In order to enhance diagnostic accuracy of breast cancer, computer aided diagnosis (CAD) systems have been built. Ultrasound is one of the most frequently used methods for early detection of breast cancer. Currently, the accuracy of CAD systems based on ultrasound images is about 90% and needs further enhancement in order to save lives of the undetected. A meaningful approach to do this is to explore new and meaningful features with discriminating ability and incorporate them into CAD systems. Recently, from a thorough investigation of the images, we extracted a new geometric feature related to the mass shape in ultrasound images called Central Regularity Degree (CRD). The CRD reflects the degree of regularity of the middle part of the mass. To demonstrate the effect of CRD on differentiating malignant from benign masses and the potential improvement to the diagnostic accuracy of breast cancer using ultrasound, this study evaluated the diagnostic accuracy of different classifiers when the CRD was added to five powerful mass features obtained from previous studies including one geometric feature: Depth-Width ratio (DW); two morphological features: shape and margin; blood flow and age. Artificial Neural Networks (ANN), K Nearest Neighbour (KNN), Nearest Centroid, Linear Discriminant Analysis (LDA), and Receiver Operating Characteristic (ROC) analysis were employed for classification and evaluation. Ninety nine breast sonograms- 46 malignant and 53 benign- were evaluated. The results reveal that CRD is an effective feature discriminating between malignant and benign cases leading to improved accuracy of diagnosis of breast cancer. The best results were obtained by ANN where the area under ROC curve (Az) for training and testing using all features except CRD was 100% and 81.8%, respectively, and 100% and 95.45% using all features. Therefore, the overall improvement by adding CRD was about 14%, a significant improvement.

Keywords: *Ultrasound, neural networks, breast cancer, early detection.*

1. INTRODUCTION

Breast cancer is the second leading cause of death after lung cancer in women all over the world (Jemal *et al*., 2005). The survival rate of breast cancer patients depends on the stage of diagnosis; patients in stage 0 (early stage) are more likely to reach cancer- free state. Therefore, early detection of breast cancer is the key to patient survival (Norman *et al*., 2006). Ultrasound as an imaging tool in medicine has been used for nearly 70 years (Kane *et al*., 2004). It is a common diagnostic medical procedure that uses inaudible sound pressure with a high frequency. The sound waves break through a medium and the echo of signals are recorded and transformed into a video or photographic image (Novelline, 1997). Sonograph in breast cancer diagnosis field has been used for differentiating solid from cyst masses. The role of ultrasound image has been expanded by improving the quality of the image and now ultrasounds have become a complementary test to mammographs for differentiating benign from malignant masses (Lee *et al*., 2008; Song *et al*., 2005). Unlike mammography, ultrasound can deal with dense breast tissue. Therefore, it is highly recommended for women with dense breast tissue or women under 40 years of age to take an ultrasound examination (Lee *et al.*, 2008). Ultrasound image is processed carefully to differentiate malignant from benign masses. The radiologist extracts a number of mass features from ultrasound image such as, shape, margin, orientation, echogenic pattern, posterior acoustic features, effect on surrounding parenchyma and Calcifications (Table 1) (Helmut, 2008; Popli, 2002). All these features are used to differentiate benign from malignant masses.

Table 1. Ultrasound mass features

Feature	Description
shape	The mass takes two main shapes, regular and irregular, the regularing masses, (or oval) are probably benign and irregular masses (speculated) are probably malignant.
margin	The border that separates the mass from the neighbour normal tissue, it can be clear or well defined which is suggestive of benign or it can be blur or ill defined which is probably malignant.
orientation	The long axis of the mass can be parallel to the skin line which is suggestive of benign.
posterior acoustic	The shadow behind the mass usually caused by malignant lesion.
Echogenic pattern	This feature reflects the internal mass Echogenicity (density of the mass).
Surrounding tissue	The effect of the mass on the surrounding tissues depends on the mass type, for example, sold mass may compact the neighbour tissues.
Classification	The presences of calcium inside the breast tissues..
Blood flow	The speed of the blood inside the mass tissues
Envelop	Most benign masses are enveloped or partially enveloped.

Several breast cancer Computer Aided Diagnosis (CAD) systems have been built upon ultrasound to differentiate benign from malignant lesions. A computerized detection and classification algorithm (Drukker *et al*., 2004) has been reported for differentiating malignant from benign masses based on ultrasound images. The study used Artificial Neural Network (ANN) and Receiver Operating Characteristic (ROC) curve for classification and evaluation, respectively, based on 400 cases for training and 458 cases for testing. The resulting area under the ROC curve (A_z values) were 0.87 with the training and 0.81 with the testing. A further study (Drukker *et al*., 2005) with the same algorithm used 609 cases obtained from two different datasets acquired from two different ultrasound platforms. The A_z values achieved by the study were between 0.8 and 0.86. Also, Artificial Neural Network (ANN) has been evaluated Song *et al*.(2005) using age and three ultrasound features (margin sharpness, intensity of absorbed sound waves by the mass margin and angular continuity of the margin). The study obtained an accuracy 0.856 ± 0.058 under ROC curve. Decision trees have been used to diagnose breast tumours using texture features with 95.5% (86/90) accuracy (Kuo et al., 2001). Furthermore, Linear Discriminate Analysis (LAD) has been used (Lee *et al.*, 2008) using six features: two of these features are geometric (compactness and orientation) and the others are echo features (intensity ratios of the regions below the two sides of a mass, intensity ratios of the regions below the mass, homogeneity, and DW ratio (depth over width). The study used ROC for evaluation and the area under the curve A_z value was 0.92.

The aim of this paper is to present a new mass feature that measures the Central Regularity Degree (CRD) of the mass and evaluate the potential improvement to the diagnostic accuracy of breast cancer using ultrasound when the new feature is added to the CAD system.

2. MATERIALS AND METHOD

This paper evaluates 99 cases; 46 are malignant and 53 are benign. All cases were obtained from The Digital Database for Breast Ultrasound Image (DDBUI) (Tian *et al*., 2008). All images were collected by the Second Affiliated Hospital of Harbin Medical University from 2002 to 2007. Each case in the database contains 1 to 6 images and a text file that lists important information of the patient and the lesions, such as, age, family history, shape, margin, size, blood flow, echo and microcalcification number and shape. All these features were taken by five experts.

2.1. Feature Extraction and Selection

Learning tasks such as classification and clustering are challenged by high dimensional data. Such data may have many noisy features which make the learning task very complex. The process of removing noisy data (irrelevant and redundant) or choosing a sub set of features (relevant) from a given set of features is called feature selection (Blum & Langley, 1997; Gilad-Bachrach *et al*., 2004). In addition to the previously mentioned features in the database, we extracted a new geometric feature related to the mass shape called Central Regularity Degree (CRD). The CRD reflects the degree of regularity of middle part of the mass. As illustrated in Figure 1, the mass boundary in this image was defined previously by experts as the thick white line (Tian et al., 2008). To find CRD, we draw, on ultrasound image Figure 1, the smallest rectangle that contains the complete mass using any image editor software. The rectangle lines X and Y represent the mass width and the mass depth, respectively. Then we divided the rectangle horizontally into three equal parts; upper, middle and lower. Next, for the middle part of the mass we find the length of the horizontal line that is parallel to the rectangle line (X) and connects the closest two points on the mass border (Z). Finally, we find the ratio of Z to the rectangle line (X) equation 1. The output value represents the Central Regularity Degree of the middle part of the mass.

$$CRD = Z \div X \quad (1)$$

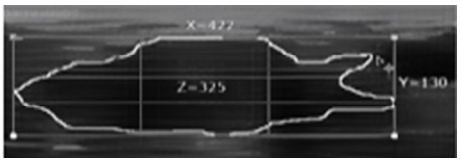

Figure. 1. X is the rectangle line parallel to skin line, Y is the rectangle width and Z is the shortest line in the middle part.

Frequencies of specific ultrasound features in both malignant and benign cases are shown in (Table 2). The feature would be considered as a good feature if it clearly separated benign from malignant cases; for example, the margin was considered as a good feature because most of malignant cases (38 out of 46) were blur and most of benign cases (38 out of 53) were clear. On the other hand, the mass echo was considered a not good feature because most of benign (34 out of 53) and malignant (42 out of 46) cases were not equable.

Table 2**.** Frequency of ultrasound features in 99 cases (46 are malignant(M) and 53 are benign(B))

Feature	M	B	Quality
Age mean	46.4	38	Good
Shape			
Regular	2	29	Good
Irregular	44	24	
Margin			
Clear	8	38	Good
Blur	38	15	
Echo			
Equable	4	19	Not good
Not equable	42	34	
Envelope			
Enveloped	8	9	
Partially	5	9	Not good
No	33	35	
Microcalcification			
Big	2	0	good
Cluster	2	1	
Needle	21	6	
None	21	46	
Blood level			
Level 1	14	41	Good
> 1	32	12	
WD ratio			
>=1.34	17	36	Good
<1.34	29	17	
CRD			
>=0.7	10	39	Good
<0.7	36	14	
Diameter			
<=1	2	8	Not
>1 and <=2	24	26	
>2	20	19	

Each case, S in the ultrasound dataset is represented as a vector of case features $S=\{S_1,...,S_n\}$. The set of features that strongly related to breast cancer is selected by using Hierarchal clustering (Johnson, 1967) and Self Organizing Map (SOM) as follows:

1- Build a state space starting from empty set in the root and add features one by one until we reach the set of all features.

2- Use sequential search starting from the root to find the node that separates benign clusters from malignant clusters by applying the following steps:

   a. Apply Hierarchal clustering.
   b. Find the best cut off point that differentiates benign clusters from malignant clusters.
   c. Compute and save the accuracy and the node index.
   d. Repeat a-d until all nodes in the state space are visited

3- Select the node with the highest accuracy.

4- Validate the results using Self Organizing Map (SOM) clustering

2.2. Classification.

The aim of any classification method is to classify objects into two or more groups based on the object attributes. There are two groups of classifiers: supervised and unsupervised classifiers. The main difference between the two groups is that, supervised methods use known output data whereas unsupervised methods rely solely on input data to find clusters (Japkowicz, 2001). This paper applies four supervised classifiers: Multilayer Feed Forward Neural Network (MFFNN), Nearest Centroid (NC), K nearest neighbour (KNN) and Linear Discriminant Analysis (LDA).

Multilayer Feed Forward Neural Network (MFFNN)

This is a supervised Neural Network used significantly in classification tasks. It contains a number of neurons, organized in layers. Every neuron in a layer is linked with all neurons in the previous layer. Each

link has a weight. The weights of the links represent the knowledge of a network. The number of layers and neurons in each layer depend on the nature of the problem. The learning phase of the network is a process by which weights of a neural network are modified to encode the knowledge of the network. The learning of the FFNN is usually done by using back propagation algorithm (Rojas, 1996) which contains two phases: The first phase is forward where the input vector is transmitted to the network layers until the output layer. At the end of this phase the algorithm calculates the error by finding the difference between the network output and the actual target. The second phase is backward where the error is back propagated from the output layer to input layer and the weights of all links are updated. In this paper, we used a Multilayer Feed forward Neural Network with 6 neurons in the input layer, one hidden layer containing 15 neurons and one neuron in the output layer. The network was trained using Scaled Conjugate Gradient back propagation algorithm (SCGBP) and Logistic function as the neuron activation function.

Optimizing the number of hidden neurons in the network is still a challenge. Insufficient number of hidden neurons results in two problems: the first is the under fitting that result from selecting small number of hidden neurons. The second problem is over fitting that result from selecting a large number of hidden neurons. To overcome these problems, this paper uses Self Organizing Map to optimize the number of hidden neurons in the network as described in Samarasinghe (2010). The algorithm starts with training a feed forward neural network with a relatively large number of neurons in the hidden layer. Then it reduces the number of hidden neurons by removing the redundant neurons that form correlated associations with other neurons. The net weighted input u_i to neuron *i* and the output y_i of each neuron in the hidden layer is given by the following equation:

$$u_i = \sum_{m=1}^{r} x_m w_{im}; \quad y_i = \frac{1}{1 + e^{-u_i}} \tag{2}$$

where x is an input vector, r is the number of inputs including bias and w_i is the weight vector between input vector x and neuron i. From equation 2 the net weighted input to the output neuron v and the final output of the neural network are:

$$v = b_0 + \sum_{i=1}^{n} y_i b_i; \quad z = \frac{1}{1 + e^{-v}} \tag{3}$$

where b_0 is the bias input to the output neuron, b is the weight vector between hidden neurons and the output neuron.

Form the above equations, the effect of hidden neurons in the classification result depends on the input-hidden layer weights and hidden-output layer weights so that we can describe each neuron in the hidden layer as:

$$Ne_i = \{w_{i0} w_{i1}, w_{i2} \cdots w_{im}, b_i\} \tag{4}$$

Now, we have a hidden neuron weights matrix NE where each row in the matrix represents one neuron. To reduce the complexity of the neural network, we will remove the redundant neurons. To do that, we applied SOM to find the distribution of hidden neuron vectors over SOM map and grouped similar neurons into clusters. The number of different clusters results from SOM indicates the optimum number of hidden neurons.

K-Nearest Neighbour (KNN).

In this classifier, the experimental samples are represented as marked points in the space where each mark denotes one class. For the new instance, the classifier represents the instance in the same space and calculates the distance between it and the experimental samples. The label of the new instance depends on the labels of the K closest points to the new instance. The instance is labelled with the class label that has largest number of points within K closest points (Wu *et al*., 2007)**.**

Nearest Centroid (NC).

In this classifier, the classification is done by calculating the mean (centroid) of each class. For the new object x, the algorithm calculates the distance between the new object and the class means and the object is labelled with the label of the closest class centroid (equation 5) (Marcoulides, 2004).

$$\forall x \in i \leftrightarrow dis(x, M_i) \leq dis(x, M_j) \quad \forall \quad i \neq j \tag{5}$$

where M_i is the mean vector of the class *i* and $dis(x,M_i)$ is the distance between the instance x and the mean vector of the class i.

Linear Discriminant Analysis (LDA).

This classifier uses covariance matrix to build a hyperplane between deferent classes by maximizing between

to within variance ratio for the classes (equation 6) (Balakrishnama & Ganapathiraju, 1998) such that.

$$P(i \mid x) > p(j \mid x) \quad \forall i \neq j \tag{6}$$

The probability of x belongs to class i is not easy to compute so the simplest mathematical formula of LDA is:

$$f_i(x_k) = \mu_i C^{-1} x_k^T - \frac{1}{2}\mu_i C^{-1} \mu_i^T + \ln(p_i) \tag{7}$$

where μ_i is the mean vector of class i, C^{-1} is the inverse of covariance matrix of the dataset and p_i is the probability of class i. x_k belongs to class i if and only if:

$$f_i(x_k) \geq f_j(x_k) \quad \forall i \neq j \tag{7}$$

2.3. Self Organizing Map (SOM)

SOM is an unsupervised neural network to represent the high-dimensional data in a low-dimensional space. Also, it is an effective and powerful tool for classification and clustering. To build SOM we first determine the topology and the number of nodes in the map. The topology of SOM layer determines the physical position of each neuron in the layer such as, a grid, hexagonal or random topology. Then we begin the process of network training as follows (Samarasinghe, 2007):

1. The weight of each node in the map is initialized.
2. Select vector from training sample randomly.
3. Find the node in the map closest to the input vector by finding the distance between the input vector and map nodes (equation 8). The closest node is usually called the Best Matching Unit (BMU):

$$\text{Dist}(x, j) = \sum_{i=0}^{n} (w_{ji} - x_i)^2 \tag{8}$$

where, w_j is weight vector of node j, x is the input vector and n the length of vector.

4. Find radius of the neighbourhood using equations 9 :

$$\sigma(t) = \sigma_0 e^{\frac{-t}{\lambda}}; \quad \lambda = \frac{NOI}{MR} \tag{9}$$

where σ_0 is the initial radius of neighbour and usually is equal the radius of the map, *t* is the iteration number, λ is the time constant, *NOI* is the total number of iterations and *MR* is the map radius

5. Any nodes found within the radius of the *BMU* are adjusted to move them closer to the input vector (equation 10).

$$w(t+1) = w(t) + \theta(t)L(t)(x(t) - w(t)); \quad L(t) = L_0 e^{\frac{-t}{\lambda}}; \quad \theta(t) = e^{(dfBMU^2 \div 2\sigma^2(t))} \tag{10}$$

where *L(t)* is the learning rate and *dfBMU* is the distance of neighbour node from BMU.

6. Repeat steps 2-5 for N iterations

3. RESULTS AND DISCUSSION

3.1. Feature selection

In this paper, the hierarchical clustering and self organizing map were used for feature selection. We started with Hierarchical clustering to find a set of features that separates benign cases and malignant cases into different clusters. The Hierarchical clustering found age, shape, margin, blood level, DW and our new feature CRD the features that strongly related to breast cancer. The dataset was divided into 9 different clusters. The distribution of malignant samples was: 39 out of 46 cases were distributed over 3 different clusters with 0.84 sensitivity (ratio of malignant cases in the 3 clusters to total malignant cases). On the other hand, 42 out of 53 benign cases were distributed over 6 clusters with 0.793 specificity (ratio of benign cases in the 6 clusters to total benign cases). The hierarchal clustering produced 81.8% accuracy. To confirm the above results, we used SOM to find the distribution of the 99 ultrasound samples over SOM map using the same features. The dataset was distributed over different regions on the SOM map where, most of malignant cases (41 out of 46) were distributed in the upper part of SOM and most benign cases (39 out of 53) were distributed in the lower part of SOM (Figure 2C). The SOM U-matrix clearly divided the upper part of SOM into three clusters that appear in the U-matrix as dark blue regions (Figure 2B). To clarify the boundary of each cluster in SOM map, we used K-mean clustering (k=9) as in hierarchal clustering) to cluster the neurons of SOM (Figure 2A). By analysing the 9 clusters we found 89% of malignant cases were distributed over 3 clusters (1, 2 and 3) and 73.5% of benign cases were distributed over the other 6 clusters. Both Hierarchical clustering and SOM found the above features strongly related to breast cancer.

3.2. Classifications

This step has been divided into two main stages: In the first stage, we applied the four classifiers; KNN, MFFNN, NC and LDA, on the dataset using all features including CRD. For KNN, firstly, we must determine the value of K, which represents the number of neighbours that controls the class label of the new instance. To do that, this paper started with large k=n down to k=1, where n is the number of experimental samples minus one. The best result was obtained when the value of k = 3. The MFFNN is more complicated than KNN. In the MFFNN, we must take into account the optimal number of neurons in the hidden layer. To do this, firstly, we trained and tested MFFNN using a large number of hidden neurons and reduce the number gradually. Every time, we compared the results with the previous results until the best results were achieved. The best results were obtained using 15 hidden neurons. Secondly, despite the goodness of the results obtained with the 15 hidden neurons, we applied SOM clustering on 15 hidden neurons used in the previous MFFNN. The purpose was to determine whether there is a network less complicated than the 15 hidden neuron network and still gives good results. As input into SOM, a hidden neuron in the hidden layer is represented by 8 attributes; weights of the 6 inputs, and weight of the bias and hidden-output weight. To build the SOM we started from selecting the SOM topology, this paper used hexagonal topology. Then, a 4x5 hexagon map was built and trained as described previously. The SOM divided the samples (hidden neurons) over 9 different clusters. According to Samarasinghe (2010), the number of different clusters represents the number of optimum hidden neurons. To verify the goodness of the 9 hidden neurons instead of 15 neurons, we built an MFFNN using 9 neurons in the hidden layer. Then, the new MFFNN was trained and tested using the same dataset. The output results obtained from 9 hidden neurons neural network were compared with the results of MFFNN using 15 hidden neurons and the values of accuracy, sensitivity and specificity were found equal in both networks. The A_z value under the ROC curve was 95.45% with 100% sensitivity and 90.9% specificity. The results show that the SOM reduced the number of hidden neurons without any effect on the classification performance and reduced the complexity of the neural networks. The outputs of different classifiers obtained from the first stage are shown in (Table 3). The MFFNN was the superior classifier with 100% sensitivity, 90.9% specificity and 95.46% A_z value under the ROC curve for the test set. NC is the worst classifier. KNN and LDA had similar A_z values but LDA was a better discriminator of malignant cases.

Table 3. The performance of different classifiers using all features. (SN is sensitivity, SP specificity and Az is the area under the ROC curve).

classifier	Training			Testing		
	SN	SP	Az	SN	SP	A_z
KNN	85.7%	92.8%	89.6%	81.9%	90.9%	86.4%
NC	82.8%	75.6%	80.5%	100%	63.6%	81.8%
MFFNN	100%	100%	100%	100%	90.9%	95.4%
LDA	82.8%	85.7%	84.4%	90.9%	81.8%	86.4%

Table 4. The performance of different classifiers using all features except CRD).

classifier	Training			Testing		
	SN	SP	Az	SN	SP	A_z
KNN	80%	92.8%	85.7%	90.9%	90.9%	90.9%
NC	82.8%	75.6%	80.5%	90.9%	63.6%	77.3%
MFFNN	100%	100%	100%	81.8%	81.8%	81.8%
LDA	80%	83.3%	81.8%	90.9%	72.7%	81.4%

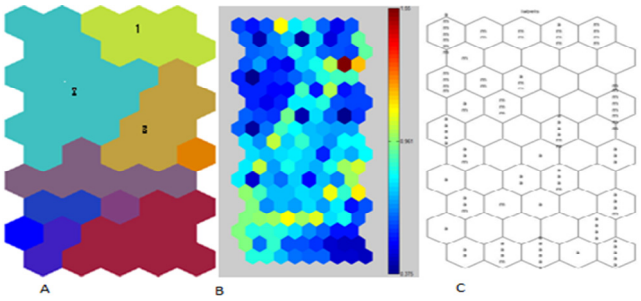

Figure 2. SOM clustering of ultrasound data A) Clusters (colour coded). B) SOM U-matrix (the distance between the neighbour nodes in the SOM lattice is represented by a colour bar that appears on the right side of the figure, the distances range from dark blue (small distance) to dark red (large distance). C) The distribution of the benign (a) and malignant (m) cases over the SOM lattice.

In the second stage, we applied the same classifiers (KNN, NC, MFFNN and LDA) on the same dataset using all features except CRD. The output results of different classifiers are shown in (Table 4). By comparing the results of deferent classifiers obtained from the first and second stages we found that: the sensitivity of KNN and LDA in the training phase were improved by 5.7% and 2.8%, respectively, by adding CRD. Also, the specificity of LDA was increased from 83.3% to 85.7% and the overall accuracies of LDA and KNN have increased. In the testing phase the sensitivity of MFFNN and NC were improved by 18.2% and 9.1%, respectively, by adding CRD. Also the specificity of MFFNN and LDA were increased from 81.8% to 90.9% and from 72.7% to 81.8%, respectively. The overall accuracy of the three classifiers, MFFNN, NC and LDA, have been enhanced by adding the CRD.

4. CONCLUSION

Early detection of breast cancer is the key to patient survival. Ultrasound has become widely used for early detection of breast cancer. To enhance the diagnosis accuracy of breast cancer, several CAD systems have been built. This paper evaluated the effect of using a new feature called Central Regularity Degree (CRD) on classification accuracy of different classifiers. The classification results of different classifiers have shown that the new feature CRD increased the performance of different CAD systems in differentiating malignant from benign lesions. Specifically, the CRD increased the overall accuracy of the best ANN classifier by 14%, a significant improvement. In future, a larger dataset will be used to confirm the effect of CRD in early detection of breast cancer.

REFERENCES

Balakrishnama, S., & Ganapathiraju, A. (1998). Linear Discriminant Analysis - A brief Tutorial: Institute for Signal and information processing http://citeseer.ist.psu.edu/contextsummary/1048862/0

Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence, 97*(1-2), 245-271. doi:10.1016/s0004-3702(97)00063-5

Drukker, K., Giger, M. L., & Metz, C. E. (2005). Robustness of computerized lesion detection and classification scheme across different breast US platforms. *Radiology, 237*(3), 834-840. Article. doi:10.1148/radiol.2373041418

Drukker, K., Giger, M. L., Vyborny, C. J., & Mendelson, E. B. (2004). Computerized detection and classification of cancer on breast ultrasound. *Academic Radiology, 11*(5), 526-535. Article.

Gilad-Bachrach, R., Navot, A., & Tishby, N. (2004). *Margin based feature selection - theory and algorithms*. Paper presented at the Proceedings of the twenty-first international conference on Machine learning.

Helmut , E. B. M. (2008). *The Practice of Breast Ultrasound Technigues.Findings.differential Diagnosis* (2 ed.): Thieme Verlag.

Japkowicz, N. (2001). Supervised versus unsupervised binary-learning by feedforward neural networks. *Machine Learning, 42*(1-2), 97-122. Article

Jemal, A., Murray, T., Ward, E., Samuels, A., Tiwari, R. C., Ghafoor, A., Feuer, E. J., & Thun, M. J. (2005). Cancer Statistics, 2005. *CA Cancer J Clin, 55*(1), 10-30. doi:10.3322/canjclin.55.1.10

Johnson, S. (1967). Hierarchical clustering schemes. *Psychometrika, 32*(3), 241-254. doi:10.1007/bf02289588

Kane, D., Grassi, W., Sturrock, R., & Balint, P. V. (2004). A brief history of musculoskeletal ultrasound: 'From bats and ships to babies and hips'. *Rheumatology, 43*(7), 931-933. doi:10.1093/rheumatology/keh004

Kuo, W. J., Chang, R. F., Chen, D. R., & Lee, C. C. (2001). Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images. *Breast Cancer Research and Treatment, 66*(1), 51-57. Article

Lee, G. N., Fukuoka, D., Ikedo, Y., Hara, T., Fujita, H., Takada, E., Endo, T., & Morita, T. (2008). Classification of benign and malignant masses in ultrasound breast image based on geometric and echo features. In E. A. Krupinski (Ed.), *Digital Mammography, Proceedings* (Vol. 5116, pp. 433-439). Berlin: Springer-Verlag Berlin.

Marcoulides, G. A. (2004). The elements of statistical learning: Data mining, inference and prediction. *Structural Equation Modeling-a Multidisciplinary Journal, 11*(1), 150-151. Book Review

Norman, S., Localio, A., Zhou, L., Weber, A., Coates, R., Malone, K., Bernstein, L., Marchbanks, P., Liff, J., Lee, N., & Nadel, M. (2006). Benefit of screening mammography in reducing the rate of late-stage breast cancer diagnoses (United States). *Cancer Causes and Control, 17*(7), 921-929. doi:10.1007/s10552-006-0029-3

Novelline. (1997). Squire's Fundamentals of Radiology In (5th ed., pp. 34–35): Harvard University Press.

Popli. (2002). Pictorial essay : Sonographic differentiation of solid breast lesions. *Indian J Radiol Imaging, 12*(2), 275-279.

Rojas, R. (1996). Neural Networks - A Systematic Introduction (pp. 151-167). Berlin: Springer-Verlag

Samarasinghe, S. (2007). *Neural networks for applied sciences and engineering : from fundamentals to complex pattern recognition* Boca Raton.

Samarasinghe, S. (2010). *neural networks for water system analysis:from fundamentals to complex pattern recognition*. Paper presented at the Hydrocomplexity:New Tools for Solving Wicked Water Problems.

Song, J. H., Venkatesh, S. S., Conant, E. F., Cary, T. W., Arger, P. H., & Sehgal, C. M. (2005). Artificial neural network to aid differentiation of malignant and benign breast masses by ultrasound imaging. *Medical Imaging 2005: Ultrasonic Imaging and Signal Processing* (Vol. 5750, pp. 148-152).

Tian, J. W., Wang, Y., Huang, J. H., Ning, C. P., Wang, H. M., Liu, Y., & Tang, X. L. (2008). *The Digital Database for Breast Ultrasound Image*. Paris: Atlantis Press.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2007). Top 10 algorithms in data mining. *Knowl. Inf. Syst., 14*(1), 1-37. doi:1327436